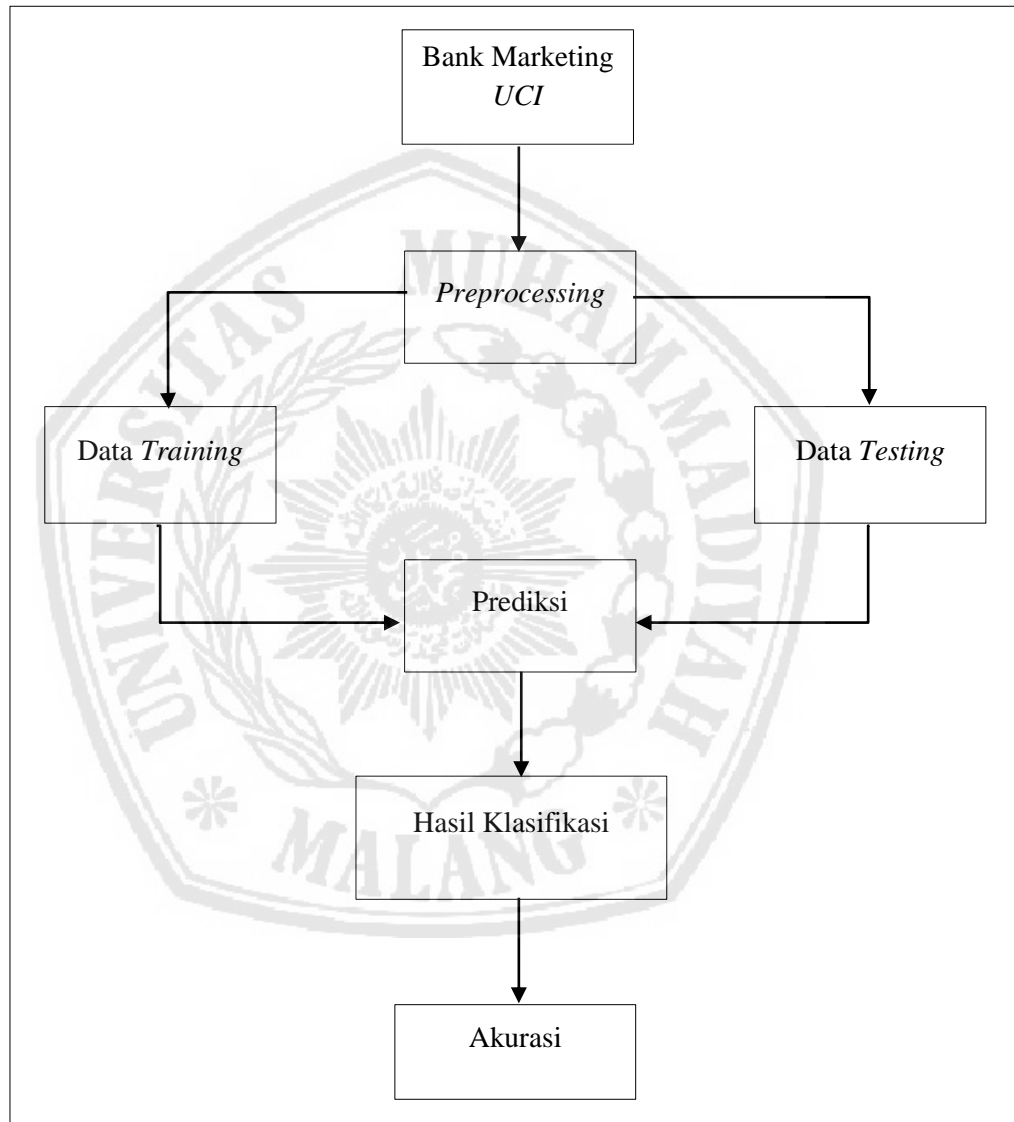


BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1. Diagram Sistem

Gambar dibawah ini merupakan diagram dari sistem secara umum :



Gambar 3.1. Diagram sistem

Dimulai dari dataset bank marketing *UCI* akan dilakukan *preprocessing* menggunakan data *transformation* terlebih dahulu. Adapun tujuan menggunakan data *transformation* agar mempermudah proses

implementasi pada program serta proses perhitungan yang dihasilkan akan lebih efisien. Setelah melalui tahap *preprocessing*, data tersebut dibagi menjadi data *training* dan data *testing* menggunakan *Cross Validation* sebanyak 10-fold *cross validation*.

Selanjutnya untuk proses prediksi akan digunakan dua metode, yaitu untuk metode pertama menggunakan *Naïve Bayes*, metode kedua menggunakan *Naïve Bayes* dengan menambahkan pembobotan *Gain Ratio*. Untuk pengukuran performansi akan dibandingkan tingkat akurasi antara *Naïve Bayes* yang menggunakan pembobotan *Gain Ratio* dengan *Naïve Bayes* tanpa menggunakan pembobotan *Gain Ratio* menggunakan *Confusion Matrix*.

3.1.1. Dataset

Dataset yang digunakan dalam tugas akhir ini adalah data Nasabah bank yang diambil dari bank marketing *UCI Machine Learning*. Data tersebut mempunyai 4521 *record* yang terdiri dari 16 atribut prediktor dan 1 atribut target. Atribut prediktor yang digunakan adalah :

1. *Age*
2. *Job*
3. *Marital*
4. *Education*
5. *Default*
6. *Balance*
7. *Housing*
8. *Loan*
9. *Contact*
10. *Day*
11. *Month*
12. *Duration*
13. *Campaign*
14. *Pdays*
15. *Previous*
16. *Poutcome*

Sedangkan untuk atribut target adalah y.

Dimana atribut *Age, Balance, Day, Duration, Campaign*, dan *Previous* merupakan data numerik. Sedangkan atribut *Job, Marital, Education, Default, Housing, Loan, Contact, Month, Poutcome* dan y merupakan data nominal.

3.1.2. Preprocessing

Tujuan dari *preprocessing* pada tugas akhir ini adalah untuk mempermudah implementasi didalam program serta proses perhitungan yang dihasilkan akan lebih efisien. *Preprocessing* yang digunakan yaitu *data transformation*. Berikut contoh penggambaran *preprocessing* dari beberapa data nasabah bank pada tugas akhir ini.

Tabel 3.1. Contoh data nasabah bank yang akan di-*preprocessing*

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
44	management	married	tertiary	no	6203	yes	yes	cellular	17	nov	58	1	188	1	failure	no
45	admin.	married	secondary	yes	1	no	no	telephone	2	feb	1028	2	207	1	other	no
58	entrepreneur	single	tertiary	yes	6162	no	yes	telephone	3	feb	112	1	-1	0	unknown	no
59	services	divorced	secondary	no	1	yes	no	cellular	28	jan	602	1	253	5	other	yes
41	blue-collar	single	secondary	yes	-386	no	yes	cellular	20	nov	477	1	-1	0	unknown	yes
51	management	married	tertiary	no	3463	no	yes	cellular	13	jul	371	2	166	3	other	yes

Pada dataset diatas terdapat beberapa atribut yang mengalami transformasi data. Atribut-atribut yang harus ditransformasi diantaranya adalah *job, marital, education, default, housing, loan, contact, month, poutcome* dan y.

Atribut tersebut ditransformasi agar mempermudah implementasi dalam program serta proses perhitungan yang dihasilkan akan lebih efisien, dan pola yang ditemukan lebih mudah untuk dipahami. Transformasi ini dilakukan secara manual. Berikut nilai-nilai atribut yang ditransformasikan :

job	data transformasi
admin.	1
unknown	2
unemployeed	3
management	4
housemaid	5
enterpreneur	6
student	7
blue-collar	8
self employed	9
retired	10
technician	11
services	12

marital	data transformasi
mearried	1
devorce	2
single	3

education	data transformasi
unknown	1
secondary	2
primary	3
teritiary	4

default	data transformasi
no	1
yes	2

housing	data transformasi
no	1
yes	2

loan	data transformasi
no	1
yes	2

contact	data transformasi
unknown	1
telephone	2
cellular	3

month	data transformasi
jan	1
feb	2
mar	3
apr	4
may	5
jun	6
jul	7
aug	8
sep	9
oct	10
nov	11
dec	12

poutcome	data transformasi
unknown	1
other	2
failure	3
success	4

y	data transformasi
no	1
yes	2

Tabel 3.2. Tabel nasabah bank yang sudah di-*preprocessing*

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
44	4	1	4	1	6203	2	2	3	17	11	58	1	188	1	3	1
45	1	1	2	2	1	1	1	2	2	2	1028	2	207	1	2	1
58	6	3	4	2	6162	1	2	2	3	2	112	1	-1	0	1	1
59	12	2	2	1	1	2	1	3	28	1	602	1	253	5	2	2
41	8	3	2	2	-386	1	2	3	20	11	477	1	-1	0	1	2
51	4	1	4	1	3463	1	2	3	13	7	371	2	166	3	2	2

3.1.3. Pembagian Data

Setelah melalui tahap *preprocessing*, tahap selanjutnya yaitu membagi data tersebut menjadi dua bagian, yaitu data *training* dan data *testing*. Dimana data *training* dan data *testing* akan dibagi menggunakan metode *Cross Validation*. Data tersebut dibagi sebanyak 10-fold *cross validation* karena 10 adalah jumlah yang tepat untuk mendapatkan estimasi yang terbaik. Setiap k-fold akan mendapatkan jumlah data yang sama.

Berikut pembagian data *training* dan data *testing* menggunakan 10-fold *cross validation* [14]:

Tabel 3.3. pembagian data *training* dan data *testing*

Data testing	Data training
Data 1	Data 2-10
Data 2	Data 1, data 3-10
Data 3	Data 1-2, data 4-10
Data 4	Data 1-3, data 5-10
Data 5	Data 1-4, data 6-10
Data 6	Data 1-5, data 7-10
Data 7	Data 1-6, data 8-10
Data 8	Data 1-7, data 9-10
Data 9	Data 1-8, data 10
Data 10	Data 1-9

Dimana :

- Jika data 1 digunakan sebagai data *testing*, maka data 2-10 adalah data *training*
- Jika data 2 digunakan sebagai data *testing*, maka data 1 dan data 3-10 adalah data *training*

- Jika data 3 digunakan sebagai data *testing*, maka data 1-2 dan data 4-10 adalah data *training*
- Jika data 4 digunakan sebagai data *testing*, maka data 1-3 dan data 5-10 adalah data *training*
- Jika data 5 digunakan sebagai data *testing*, maka data 1-4 dan data 6-10 adalah data *training*
- Jika data 6 digunakan sebagai data *testing*, maka data 1-5 dan data 7-10 adalah data *training*
- Jika data 7 digunakan sebagai data *testing*, maka data 1-6 dan data 8-10 adalah data *training*
- Jika data 8 digunakan sebagai data *testing*, maka data 1-7 dan data 9-10 adalah data *training*
- Jika data 9 digunakan sebagai data *testing*, maka data 1-8 dan data 10 adalah data *training*
- Jika data 10 digunakan sebagai data *testing*, maka data 1-9 adalah data *training*

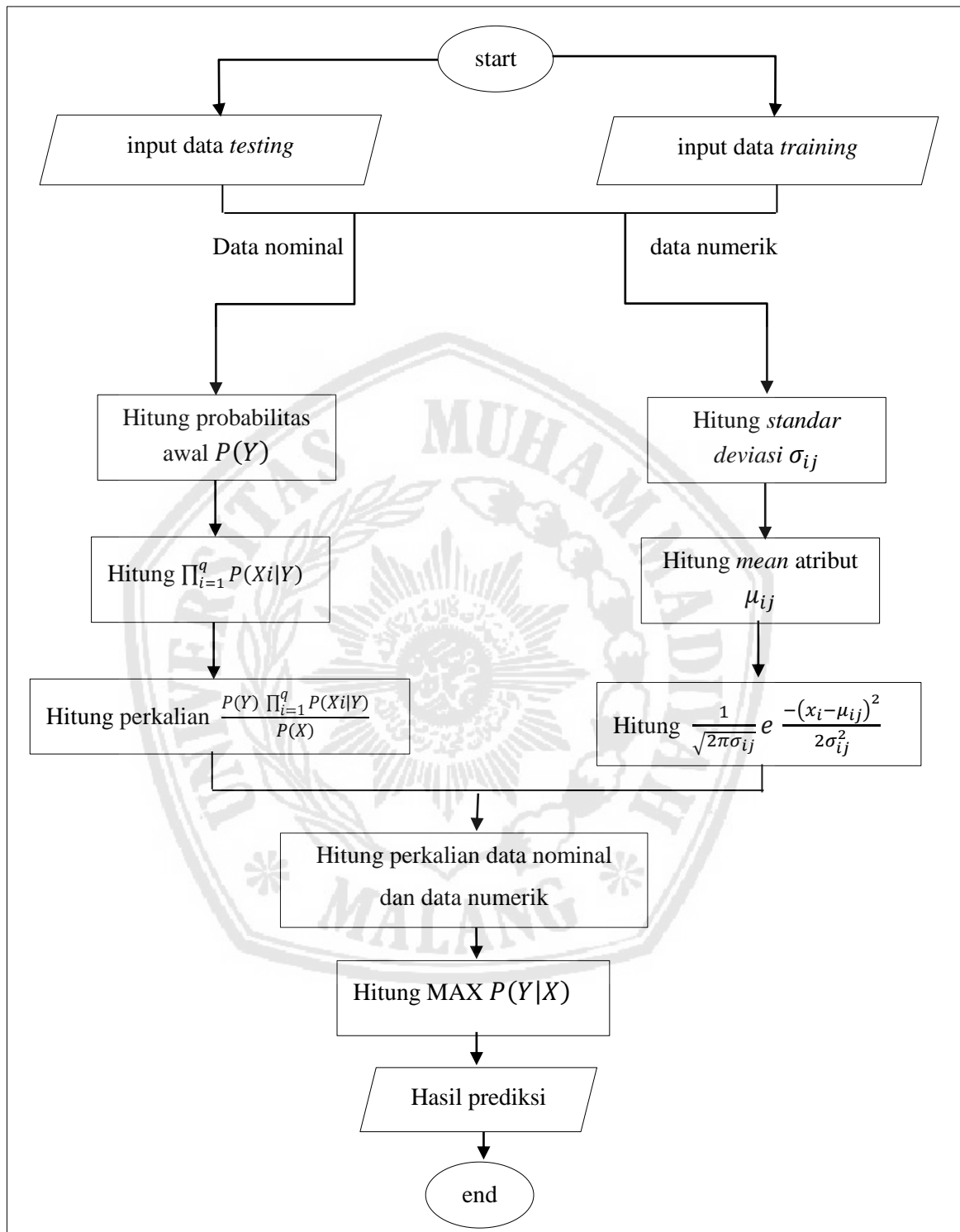
3.1.4. Prediksi

Setelah *dataset* dibagi menjadi data *training* dan data *testing*, selanjutnya *dataset* tersebut akan diprediksi hasilnya menggunakan metode yang terdapat pada tugas akhir ini. Adapun tujuan dari proses perhitungan tersebut adalah agar didapatkannya hasil prediksi dari *dataset* yang ditanyakan. Terdapat dua metode yang digunakan pada tugas akhir ini, untuk metode pertama yaitu *Naïve Bayes*, dan untuk metode yang kedua yaitu *Naïve Bayes* dengan pembobotan *Gain Ratio*:

3.1.4.1. Naïve Bayes

Pada tugas akhir ini terdapat dua cara untuk melakukan proses perhitungan menggunakan metode *Naïve Bayes*, untuk data nominal maka perhitungannya menggunakan formulasi dari persamaan (2.1) sedangkan untuk data numerik maka perhitungannya menggunakan persamaan (2.2).

Untuk proses perhitungan *Naïve Bayes* dapat dilihat pada *flowchart* berikut :



Gambar 3.2. *Flowchart* perhitungan *Naïve Bayes*

Dari gambar diatas, terdapat beberapa tahapan untuk melakukan perhitungan menggunakan *Naïve Bayes*, sebagai berikut:

1. Tahap pertama yaitu input data *testing* dan data *training*
2. Tahap kedua menghitung probabilitas, namun apabila data nominal maka :
 - a. Hitung probabilitas awal dari $P(Y)$, yaitu menghitung banyaknya jumlah data dari atribut yang ditanyakan terhadap kelas yang ditanyakan.
 - b. Hitung probabilitas independen kelas Y dari semua kelas X , yaitu menghitung banyaknya jumlah atribut yang ditanyakan terhadap semua kelas dari X
 - c. Hitung hasil dari perhitungan (a) lalu dikalikan dengan (b)

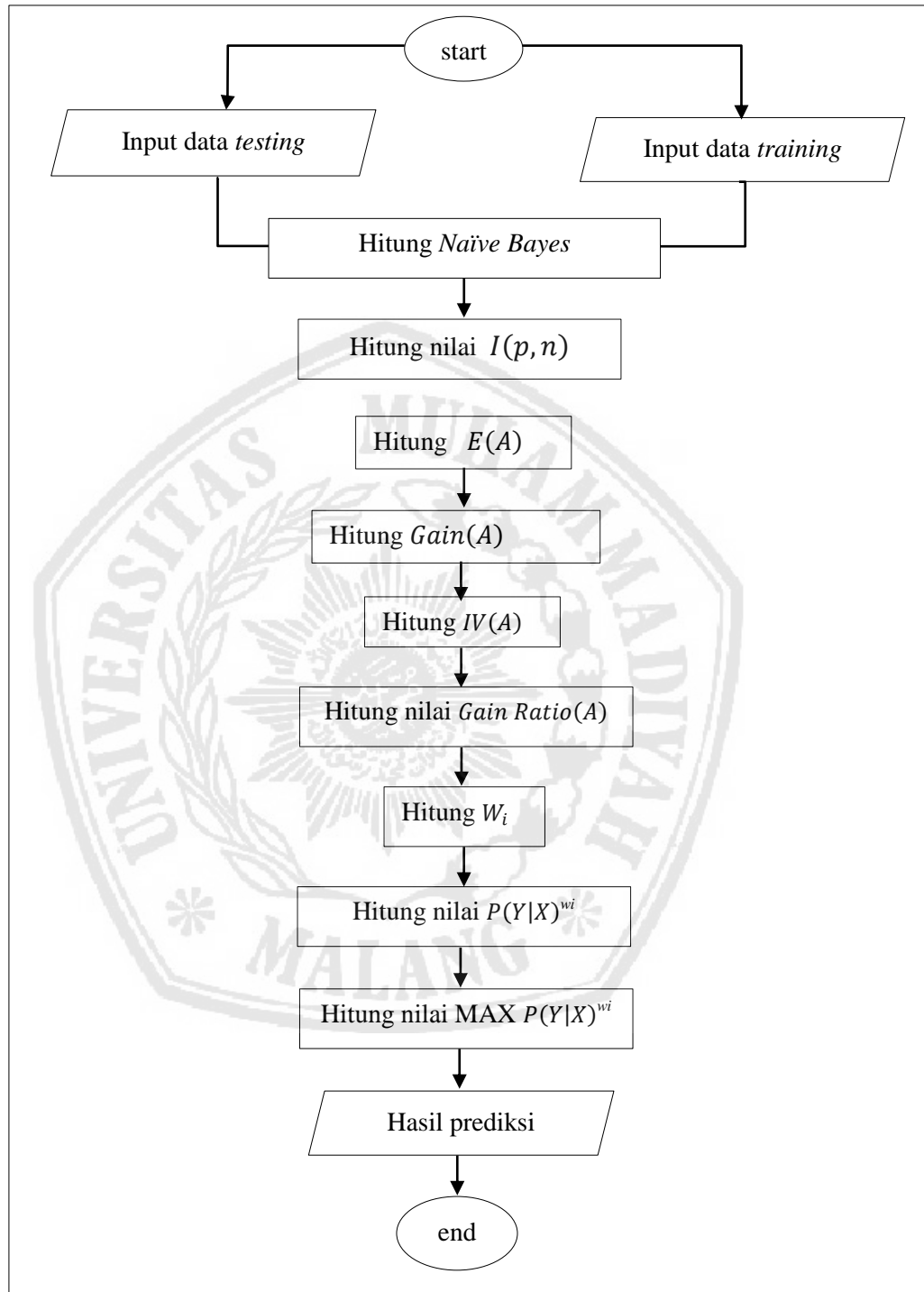
Jika data numerik maka :

- a. Cari nilai *standar deviasi* dari masing-masing parameter yang merupakan data numerik
 - b. Hitung nilai *mean*
 - c. Hitung hasil perkalian dari (a) lalu dikalikan dengan hasil (b)
3. Tahap ketiga yaitu menghitung perkalian antara hasil dari data numerik dikalikan dengan hasil data nominal
4. Tahap keempat yaitu membandingkan hasil MAX dari hasil perhitungan pada tahap 3. Nilai MAX yang didapatkan akan dijadikan sebagai hasil perhitungan menggunakan *Naïve Bayes*
5. Tahap kelima yaitu mendapatkan hasil prediksi

3.1.4.2. *Naïve Bayes* dengan pembobotan *Gain Ratio*

Metode yang kedua pada tugas akhir ini adalah *Naïve Bayes* dengan menambahkan pembobotan *Gain Ratio* untuk proses perhitungannya. Adapun tujuan menambahkan pembobotan dalam perhitungannya, karena menurut *Ferreira*, *Denison*, dan *Hand* pembobotan atribut dapat meningkatkan pengaruh prediksi.

Berikut *flowchart* untuk proses perhitungan dengan menambahkan pembobotan *Gain Ratio*.



Gambar 3.3. *Flowchart Naïve Bayes dengan pembobotan Gain Ratio*

Perhitungan pada *Naïve Bayes* dengan pembobotan *Gain Ratio* terdiri atas 10 tahapan, sebagai berikut :

1. Tahapan pertama yaitu input data *testing* dan data *training*
2. Tahapan kedua melakukan perhitungan menggunakan *Naïve Bayes*.
3. Tahap ketiga menghitung nilai p ,
4. Tahap keempat menghitung nilai $H(A)$, yaitu menghitung nilai *entropy* dari setiap atribut
5. Tahap kelima menghitung $Gain(A)$ dengan cara mengurangi hasil dari perhitungan tahap 2 dengan hasil perhitungan tahap 3.
6. Tahap keenam menghitung nilai $IV(A)$ dari setiap atribut
7. Tahap ketujuh menghitung $Gain Ratio(A)$ dengan cara membagi hasil dari perhitungan tahap 5 dengan hasil perhitungan tahap 6.
8. Tahap kedelapan menghitung w_i (pembobotan) yaitu menjumlahkan semua nilai *Gain Ratio* dari setiap atribut yang ditanyakan kemudian dibagi dengan banyaknya hasil perhitungan *Gain Ratio* dari semua atribut.
9. Tahap kesembilan yaitu menghitung *Naïve Bayes* dengan memangkatkan pembobotan *Gain Ratio*
10. Tahap kesepuluh yaitu mencari nilai MAX dari hasil perhitungan probabilitas *Naïve Bayes* yang sudah dipangkatkan dengan pembobotan

3.1.5. Akurasi

Pada tugas akhir ini bentuk pengujian yang dilakukan adalah menghitung tingkat akurasi dari metode yang digunakan yaitu *Naïve Bayes* dengan menggunakan pembobotan *Gain Ratio* dengan *Naïve Bayes* tanpa menggunakan pembobotan *Gain Ratio* menggunakan *Confusion Matrix*. Perhitungan akurasi menggunakan persamaan (2.10)

Berikut merupakan contoh perhitungan akurasi menggunakan *Confusion Matrix* :

Tabel 3.4. contoh data hasil pengujian

No	Input	Hasil tanpa sistem	Hasil sistem	Ket
1	Data 1	1	1	Benar
2	Data 2	1	1	Benar
3	Data 3	1	0	Salah

Dari tabel 3.4 maka didapatkan sebuah *Confusion Matrix* yang merepresentasikan informasi mengenai hasil menggunakan sistem dan hasil tanpa sistem. Tabel perhitungan *Confusion Matrix* dari tabel 3.4 ditunjukkan pada tabel 3.5 berikut ini :

Tabel 3.5. perhitungan *Confusion Matrix*

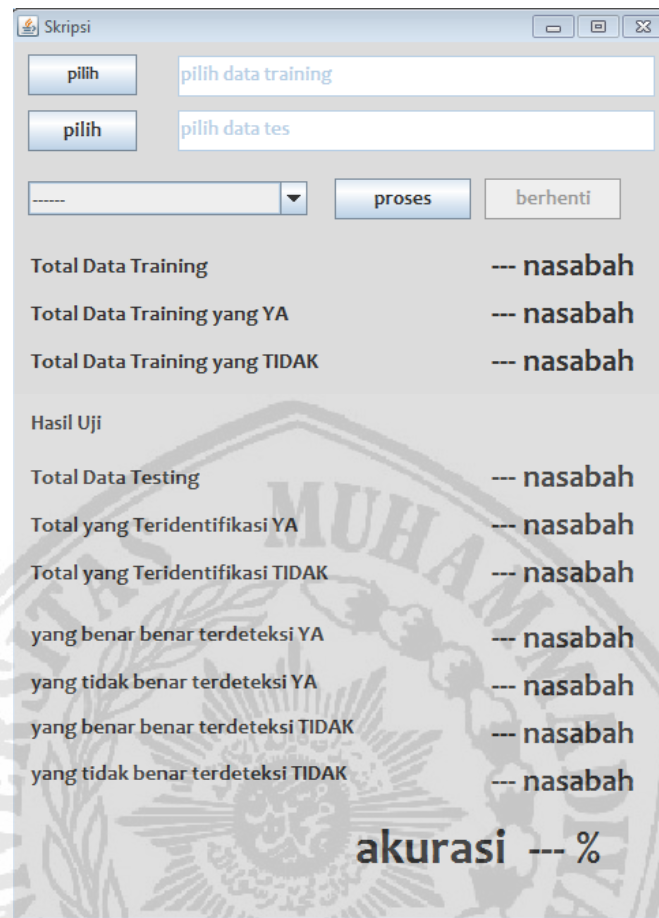
		Prediksi	
		1	0
Actual (sebenarnya)	1	TP = 2	FP = 1
	0	FN = 0	TN = 0

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + FP + TN + FN} \\
 &= \frac{2 + 0}{2 + 1 + 0 + 0} \\
 &= \frac{2}{3} = 0,66 \times 100\% = 66,6\%
 \end{aligned}$$

Dari perhitungan diatas didapatkan hasil akurasi sebesar 66,6%

3.2. Perancangan Sistem

Gambar dibawah ini merupakan *interface* sistem tugas akhir :



Gambar 3.4. Rancangan *interface* sistem

Pada gambar diatas terdapat beberapa fungsi yang digunakan pada program seperti penjelasan berikut :

- Tombol pilih data *training*, digunakan untuk memilih file data *training* yang akan di proses
- Tombol pilih data *testing*, digunakan untuk memilih file data testing yang akan di proses
- Tombol pilih metode, digunakan untuk memilih metode yang akan digunakan untuk proses perhitungan
- Tombol proses, digunakan untuk melakukan perhitungan dalam sistem menggunakan metode yang telah dipilih

Tombol berhenti, digunakan untuk mengembalikan nilai awal dari tampilan program